PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Pharmacologic and Nonpharmacologic Treatments for Major Depressive Disorder: Review of Systematic Reviews
AUTHORS	Gartlehner, Gerald; Wagner, Gernot; Matyas, Nina; Titscher, Viktoria; Greimel, Judith; Lux, Linda; Gaynes, Bradley; Viswanathan, Meera; Patel, Sheila; Lohr, Kathleen

VERSION 1 - REVIEW

REVIEWER	Klaus Linde Institute of General Practice, Technical University Munich, Germany
REVIEW RETURNED	22-Nov-2016

GENERAL COMMENTS	This manuscript reports a carefully planned and performed review of systematic reviews of pharmacologic and nonpharmacologic treatments for patients suffering from a major depressive disorder. The team has excellent experience in the clinical subject and the methods. The reporting is clear and accurate, the conclusions seem reasonable. Given the amount, the complexity, diversity and limitations of the material under review (of the original reviews) it is inevitable that there are some problems, and that reporting cannot provide detailed insight into each detail. In my view, the manuscript could be published with minor changes – in spite of some concerns. My main comment regards the selection and classification of comparators. First, while I fully understand (based on own desperate review experience) the problems the authors have with usual care controls (which are very badly described in most trials), unstructured usual care is a typical treatment option in real life. Here the authors act “scientifically clean”. Second, lumping placebo, sham interventions, waiting list and no care into a single comparator is, in my view, problematic and potentially misleading. Here the authors are very “pragmatic”. This approach does not seem fully consistent to me. Yet I understand the authors – somehow this difficult material has to be coped with… The second problem is addressed relatively well in the paper. But your argument on page 22, lines 51-53 on usual care does not convince me at all – elsewhere you implicitly argue that your “inactive” treatments seem to have different activity! Where is the need or the rationale to lump usual care with placebo, and placebo with no treatment? Finally, effect sizes in usually care controlled trials are often between those of sham- or placebo-controlled trials and waiting-list controlled trials. I can live with your strategy on a pragmatic level, but you should consider modifying your discussion of these issues. In my view the main rationale is “keep it simple” as, anyhow, there is no perfect solution. Another important question is to what extent the included reviews really cover all relevant trials. For example, our review included by the authors (ref. 36) was restricted to trials performed in primary care – excluding the vast majority of antidepressant trials. Compared

	from what I have seen the number of trials seems relatively small to me for some treatments, but I might be wrong. The authors should consider discussing in some more detail whether the included reviews really cover the trials relevant to their question comprehensively. Minor points - page 10: what means “general efficacy”? Why not simply saying “comparison with …”? Furthermore, why different outcomes were used for the two comparisons? That seems somewhat strange to me. - If a review was “superseded” you did not check for eligible trials not included in the review included by you, correct? - Page 17, line 41: for the 24 unpublished trials you refer to reference 46 – this is a review comparing pharmacologic and non-pharmacologic treatments. Is this correct? What about the unpublished placebo-controlled trials? - Just a comment: Personally, I have problems rating trials of CBT vs. SGA low risk of bias if only the person performing the interview is blinded. Would we accept this as low risk in a drug trial?

REVIEWER	Sarah Hetrick Oregon, The National Centre of Excellence in Youth Mental Health and Centre for Youth Mental Health, University of Melbourne, Australia I am an editor for the Cochrane Common Mental Disorders Group and have undertaken reviews on interventions for youth depression.
REVIEW RETURNED	24-Dec-2016

GENERAL COMMENTS	Overall, this review has been well executed, and is well written providing some interesting results. I do have some concerns about how well it represents the extant literature base for all of these interventions given it does not include all relevant trials (e.g. some trials are excluded based on them not being included in the most recent review; some are missing due to the type of comparison used). In particular I wonder how well it compares (methodologically) to a network meta-analysis that includes all relevant trials. I think the methods for reviews of reviews are still in their infancy, and there are significant limitations in terms of combining trial level evidence with meta-analytic level evidence. Granted some of these concerns, and some raised below are cited in the limitation section. Further, I think it does present a very simplistic and easy to understand summary of what the evidence base currently tells us. It might be that some cautionary language around the results (as appears in the well written and appropriately cautious and sensible later sections of the discussion) should feature in the key findings section of the discussion and in the abstract. The following are more specific comments. It is a minor point only; but in the introduction I would be interested to know how the rates of MDD compare to the worldwide rate of depression overall. The authors provide rates of MDD in Europe, but rates for depression overall as worldwide stats, meaning it is difficult to make sense of what proportion of depression is accounted for by MDD.

	The wording of the first sentence of the section “Populations etc…” (pg 6, line 53-55) is a little awkward; I wonder if the word “criteria” should be easier in the sentence e.g. The “criteria for inclusion in the review for population etc.…” More information about how a systematic review was defined would be useful for the reader e.g. did the review simply have to call itself a systematic review or did the authors have some more stringent criteria? The screening process isn’t entirely clear – did everything that was retrieved get examined at the ‘full text’ level or was there a two stage process where titles and abstracts were reviewed and then possible and eligible full text articles were retrieved and examined for inclusion? Authors need to provide detail on their criteria for determining that a review was ‘low’, ‘medium’ etc… according to the AMSTAR criteria. I am unclear about the justification for excluding ‘treatment as usual’ as a comparison group. In many ways these are more meaningful comparisons than waitlist or no treatment, against which almost any kind of intervention will show an effect. There is some evidence that waitlist control artificially increases the treatment effect. Also – a placebo control group should not be grouped with waitlist and no treatment in my opinion as the trial participant will experience this type of control in a very different way to waitlist or no treatment. It isn’t clear how the authors dealt with synthesis of individual studies with meta-analytic results? Did authors implement meta-analytic techniques similar to those used by the review author? This also raises the question about how authors synthesised results across reviews if different meta-analytic techniques were used – was this an issue? This is also an issue for risk of bias; I’m not sure relying on the risk of bias ratings of original review authors is necessarily a robust method. Cochrane guidance on meta-analyses suggests that you need to decide a-priori which analytic methods to use (fixed vs random) rather than using both. Further regarding analysis, authors should specify their sensitivity analyses. Page 11, line 53 – typo: you need a space in between the word ‘citations’ and 2042. In the next line there is a full stop that appears before the citations rather than after. I was concerned that I had misinterpreted Table 2 because it did not appear to me that all of the trials I would have expected of SGA’s to be included appeared in this table? From this table it appears they have been vastly under represented, but I wonder if that means that the authors need more explanation of their inclusion and exclusion criteria and justification of this for the reader. Within the reporting of results in the text i.e. when SMD and RR are presented, the authors should present the number of trials and number of participants on which the result is based. It is encouraging to see the comment on the issues of publication bias when reporting the results for SGA’s. Often these trials rate quite well in terms of risk of bias, but there are a number of bias’s

	outside of e.g. the Cochrane risk of bias tool that impact on pharmacological trials. Some more consistency in terms of reporting results is required; I noted that for some results the strength of evidence is reported but e.g. for non pharmacological interventions it is not. Page 19, line 17 – typo – should read “bias”. Page 23, line14 – typo: decision making. Author should also include a reference with regard to their statement about shared decision making improving treatment adherence etc…. I think the other area for future research is around efficacy of non pharmacological interventions?

REVIEWER	Andrea Cipriani Department of Psychiatry, University of Oxford, UK
REVIEW RETURNED	21-Jan-2017

GENERAL COMMENTS	This paper is heavily based on the previous work of the authors in the same field (for instance, see refs 46, 48 and 52 of the present manuscript). I think it is interesting for clinicians to have a "short" summary of all the previous evidence/papers, but the authors should make clear (since from the introduction) why the present manuscript is different from their previous publications (refs 46, 48 and 52 should probably be quoted in the introduction). I think this article could benefit from a slightly different angle, in order to avoid the replication of findings that have been already presented/discussed in the literature very recently (I suggest a larger time window and a more recent update of the search strategy). Here below a few (mainly minor) comments: The conclusion paragraph int he abstract mentioned only non-pharmacological interventions, while the paper is also about pharmacological treatments. Please amend accordingly The first sentence of the introduction is not relevant, as the article focused on MDD. I would remove it Page 6, line 56: probably the acronym PICOTS is useless (it is reported once more in the table only and not in the text) Please clarify why the search is limited between 1 January 2011 and 23 February 2016? If reference 25 is true, the authors should now include papers from 2012! I would not be so strict, but I think the search should be a bit broader and, if it is not possible to go back before 2011, I would certainly update it up to Jan 2017 (it was carried out 11 months ago, so it would be important that the authors could check whether any relevant papers have been published since Feb last year) Sorry, I couldn't find in the text, tables or appendices any definition of "systematic review" or "meta-analysis". If I am not mistaken, it would be useful for the readers to know exactly the kind of studies included in the present article (sometimes "narrative reviews" are called "systematic review" in the title, but I'm sure the authors

	wanted to include only properly carried out systematic reviews) Table 1 is not clear to me. Why agomelatine is alone? Is it because of a different mechanism of action? If so, why did you include vortioxetine among other antidepressants? I think an explanation of the rationale for these choices should be presented/discussed in the main text (or in the table itself). Table 1: sorry, what does it mean exactly "Pharmacologic interventions (for comparison with inactive interventions)"? Table 1: "We did not include combination treatments" and "as an initial monotherapy" are redundant. I would use one or the other Table 1: please provide a clear definition of MDD (which criteria used) and response (which threshold, if any). Did the authors consider change of depression scores based only on validated scale? Did they consider both self-rating and clinician-based scales? Reference 51 should be replaced by a more recent publication with updated results: Furukawa et al. Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies. Lancet Psychiatry. 2016 Nov;3(11):1059-1066. Page 21, line 5: Is "risk of adverse events" one of the outcomes? This is not a trivial issue to clarify, not only because the risk of adverse events was not listed as an outcome in the method section of this paper and also because ref 46 included non-randomised evidence (which is in conflict with the methods as stated on page 7, first two lines. Page 23, line 6: please change "We believe that our results have important clinical implications." into "We believe that our results MAY have important clinical implications."

REVIEWER	Tolulope Sajobi University of Calgary, Canada
REVIEW RETURNED	25-Jan-2017

GENERAL COMMENTS	This study investigate reviews research evidence from more than 140 pharmacologic and on-pharmacologic treatment options for major depressive disorder (MDD) obtained from systematic reviews of randomized controlled trials in adult patients with acute-phase MDD. The review resulted in identification of 15 systematic reviews on 27 comparisons of interest. This review suggested strong evidence supporting the small benefits of second-generation antidepressants but also significant rate of discontinuation due to adverse events. Cognitive behavioral therapy was found to show reliable evidence similar to those of second-generation antidepressants. There was no evidence supporting majority of non-pharmacologic interventions for treating MDD. While this study addresses an important research question, I have some methodological concerns about the pooling of evidence from the include articles. 1. On pages 10 -11, the authors provided details about meta-

	analysis of evidence. One feature of meta-analysis of evidence from multiple systematic reviews is the increase layer of sources of variations introduced. Ideally, pooling evidence from multiple systematic reviews leads to multiple sources of variations such as between-systematic review variation, within-systematic review variation, and individual-level study variation. It is not clear if the authors accounted for between-study variations in their analyses and whether the conclusions would have different have they accounted for these sources of variations. 2. It appears the authors used conventional meta-regression methods to pool evidence but use network meta-analysis in others. This is not clearly stated anywhere in the manuscript. In fact, network meta-analysis was only mentioned in the Result section. a. Please provide a detailed description of the network meta-analysis employed as part of the statistical analysis section of the manuscript. b. The authors used conventional meta-regression approaches and network meta-analytic methods to address the research questions. I would recommend that the author indicate these in the methods section of the abstract. 3. As part of the discussions about the limitations of this review, the authors need to highlight the small number of studies (low power, low sample size) as a limitation of the meta-analysis approaches investigated 4. On page 22 line 8, the authors stated “Although most of the reviews 9 had few problems in methods, conceivably these authors did miss some RCTs. Conceivably, 10 RCTs are available…”. Conceivably used twice, please revise.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

• My main comment regards the selection and classification of comparators. First, while I fully understand (based on own desperate review experience) the problems the authors have with usual care controls (which are very badly described in most trials), unstructured usual care is a typical treatment option in real life. Here the authors act “scientifically clean”. Second, lumping placebo, sham interventions, waiting list and no care into a single comparator is, in my view, problematic and potentially misleading. Here the authors are very “pragmatic”. This approach does not seem fully consistent to me. Yet I understand the authors – somehow this difficult material has to be coped with… The second problem is addressed relatively well in the paper. But your argument on page 22, lines 51-53 on usual care does not convince me at all – elsewhere you implicitly argue that your “inactive” treatments seem to have different activity! Where is the need or the rationale to lump usual care with placebo, and placebo with no treatment?

• Finally, effect sizes in usually care controlled trials are often between those of sham- or placebo-controlled trials and waiting-list controlled trials. I can live with your strategy on a pragmatic level, but you should consider modifying your discussion of these issues. In my view the main rationale is “keep it simple” as, anyhow, there is no perfect solution.

Response: We agree with the reviewer that treatment as usual reflects real-world practice. Across many systematic reviews, however, these treatments become a heterogeneous set of interventions that are often impossible to categorize because they are poorly described. In addition, in some

systematic reviews, authors equated treatement as usual with inactive treatment, an approach with which we disagree. Our argument on page 22 that usual care cannot be viewed as inactive in most cases was based on the rationale that we did not want to view usual care as inadequate care. Particularly in trials involving complementary and alternative therapies, usual care often meant drug treatment. We agree with the reviewer that this raises issues of applicability to real-life situations. Nevertheless, we think this decision was necessary to keep an already very complex review interpretable.

We did not combine placebo, sham, and waitlist comparators in any of our own quantitative analyses. We included reviews, however, that allowed these control groups. The majority of reviews used placebo-pill as a control intervention. In three instances (exercise, Third wave CBT, and psychodynamic interventions), eligible reviews used waitlist as a control. We highlight in the text that waitlist in these reviews might lead to a greater treatment effect. On page 18, we specifically caution readers that the types of inactive comparators varied and involved different magnitudes of placebo effects. Consequently, comparisons of treatment effects across different interventions have to be made cautiously.

• Another important question is to what extent the included reviews really cover all relevant trials. For example, our review included by the authors (ref. 36) was restricted to trials performed in primary care – excluding the vast majority of antidepressant trials. Compared from what I have seen the number of trials seems relatively small to me for some treatments, but I might be wrong. The authors should consider discussing in some more detail whether the included reviews really cover the trials relevant to their question comprehensively.

Response: It is indeed a main limitation of any review of systematic reviews that conclusions are limited to the interventions and scope that the source systematic reviews had assessed. In the revised Discussion, we point this out. We also emphasize that the absence of systematic reviews cannot be equated with an absence of RCTs.

• Minor points

♣ page 10: what means “general efficacy”? Why not simply saying “comparison with …”? Furthermore, why different outcomes were used for the two comparisons? That seems somewhat strange to me.

Response: General efficacy refers to the efficacy of a treatment compared with an inactive control. The term is commonly used in North America, but if the editors feel that it is inappropriate for BMJ Open, we’d be happy to replace this term.

- If a review was “superseded” you did not check for eligible trials not included in the review included by you, correct?

Response: No, we did not cross-check trials between systematic reviews that we included and those that were superseded by a more recent review. Our assumption was that a well-conducted systematic review (we rated the methodological validity of all reviews) will have a high likelihood to detect all relevant trials and will probably have searched reference lists of previous systematic reviews on the same topic.

- Page 17, line 41: for the 24 unpublished trials you refer to reference 46 – this is a review comparing pharmacologic and non-pharmacologic treatments. Is this correct? What about the unpublished placebo-controlled trials?

Response: This citation is actually correct. The point that we are trying to make here is that despite the 24 unpublished trials in this dataset, we can assume that many more unpublished studies exist that are not in this dataset.

- Just a comment: Personally, I have problems rating trials of CBT vs. SGA low risk of bias on performing the interview is blinded. Would we accept this as low risk in a drug trial?

Response: This is a difficult and controverial issue. We agree with the reviewer that we would not accept such a trial as low risk of bias if any of the arms was inactive and patients knew about their assignment to the inactive arm. This would lead to a high risk of performance bias. In a situation, however, in which two active treatments are compared, performance bias might be less of an issue. In a well-conducted RCT, patients who would have preferred the other treatment should be equally distributed across both arms. Such a trial is, of course, by far not as ideal as a double-blinded RCT, but double-blinding in trials with psychological interventions is not possible.

Reviewer 2
• Overall, this review has been well executed, and is well written providing some interesting results. I do have some concerns about how well it represents the extant literature base for all of these interventions given it does not include all relevant trials (e.g. some trials are excluded based on them not being included in the most recent review; some are missing due to the type of comparison used). In particular I wonder how well it compares (methodologically) to a network meta-analysis that includes all relevant trials. I think the methods for reviews of reviews are still in their infancy, and there are significant limitations in terms of combining trial level evidence with meta-analytic level evidence. Granted some of these concerns, and some raised below are cited in the limitation section. Further, I think it does present a very simplistic and easy to understand summary of what the evidence base currently tells us. It might be that some cautionary language around the results (as appears in the well written and appropriately cautious and sensible later sections of the discussion) should feature in the key findings section of the discussion and in the abstract.

Response: A limitation of any review of reviews is that it has to rely on the topics that published systematic reviews covered. In the revised Discussion we emphasize the point that the absence of systematic reviews cannot be equated with an absence of RCTs. The advantage of our review is that it covers more than 140 interventions that could rarely be taken on in a single systematic review focusing on RCTs.

• The following are more specific comments.

• It is a minor point only; but in the introduction I would be interested to know how the rates of MDD compare to the worldwide rate of depression overall. The authors provide rates of MDD in Europe, but rates for depression overall as worldwide stats, meaning it is difficult to make sense of what proportion of depression is accounted for by MDD.

Response: In the revised manuscript, based on another reviewer's comment, we deleted the first sentence and focus just on MDD.

• The wording of the first sentence of the section "Populations etc…" (pg 6, line 53-55) is a little awkward; I wonder if the word "criteria" should be easier in the sentence e.g. The "criteria for inclusion in the review for population etc…."

Response: We revised this sentence.

• More information about how a systematic review was defined would be useful for the reader e.g. did

the review simply have to call itself a systematic review or did the authors have some more stringent criteria?

Response: We used a definition based on the Cochrane handbook. In the revised manuscript, we added the definition as a footnote to Table 2.

• The screening process isn't entirely clear – did everything that was retrieved get examined at the 'full text' level or was there a two stage process where titles and abstracts were reviewed and then possible and eligible full text articles were retrieved and examined for inclusion?

Response: It was a 2-stage process with a dual review of abstracts and full texts. We tried to present this process more clearly in the revised paper.

• Authors need to provide detail on their criteria for determining that a review was 'low', 'medium' etc… according to the AMSTAR criteria.

Response: In the revised manuscript, we present AMSTAR ratings for all included studies as Supplementary Files 3.

• I am unclear about the justification for excluding 'treatment as usual' as a comparison group. In many ways these are more meaningful comparisons than waitlist or no treatment, against which almost any kind of intervention will show an effect.

Response: We agree with the reviewer, treatment as usual is a meaningful comparison that reflects real-world practice. Across many systematic reviews, however, these treatments become a heterogeneous set of interventions that are often impossible to categorize because they are poorly described. In addition, in some systematic reviews, authors equated treatement as usual with inactive treatment which we disagree with. Because of these challenges, we excluded usual care as a control intervention. We agree with the reviewer that this raises issues of applicability to real-life situations. Nevertheless, we think this decision was necessary to keep an already very complex review interpretable. Treatment as usual cannot be viewed as an inactive intervention in most cases. Particularly in trials involving complementary and alternative therapies, usual care often meant drug treatment.

• There is some evidence that waitlist control artificially increases the treatment effect. Also – a placebo control group should not be grouped with waitlist and no treatment in my opinion as the trial participant will experience this type of control in a very different way to waitlist or no treatment.

Response: We did not combine these different control groups in any of our analyses. We included reviews, however, that allowed these control groups. The majority of reviews used placebo-pill as a control intervention. In three instances (exercise, Third wave CBT, and psychodynamic interventions) reviews used waitlist as a control. We highlight in the text that waitlist in these reviews might lead to a greater treatment effect. On page 18, we specicifically caution readers that the types of inactive comparators varied and involved different magnitudes of placebo effects. Consequently, comparisons of treatment effects across different interventions have to be made cautiously.

• It isn't clear how the authors dealt with synthesis of individual studies with meta-analytic results? Did authors implement meta-analytic techniques similar to those used by the review author? This also raises the question about how authors synthesised results across reviews if different meta-analytic techniques were used – was this an issue? This is also an issue for risk of bias; I'm not sure relying on the risk of bias ratings of original review authors is necessarily a robust method.

Response: We added text to the Methods to make this clearer. If the original systematic reviews meta-analyzed studies that met our eligibility criteria, we present results of this meta-analysis. If the original systematic reviews combined studies that did not all meet our eligibility criteria (e.g., studies using treatment as usual as control groups), we conducted our own meta-analyses on the subset of trials that had met our inclusion criteria. We did not, however, conduct any meta-analyses across systematic reviews.

• Cochrane guidance on meta-analyses suggests that you need to decide a-priori which analytic methods to use (fixed vs random) rather than using both. Further regarding analysis, authors should specify their sensitivity analyses.

Response: To our knowledge the Cochrane handbook recommends to choose the model based on the underlying clinical heterogeneity, which we did for our analyses. We report only random effects results; fixed effects analyses can be viewed as sensitivity analyses because they give more weight to larger studies.

• Page 11, line 53 – typo: you need a space in between the word citations' and 2042. In the next line there is a full stop that appears before the citations rather than after.

Response: Thanks for pointing this out. We fixed the typo.

• I was concerned that I had misinterpreted Table 2 because it did not appear to me that all of the trials I would have expected of SGA's to be included appeared in this table? From this table it appears they have been vastly under represented, but I wonder if that means that the authors need more explanation of their inclusion and exclusion criteria and justification of this for the reader.

Response: We think that this indeed might be a slight misinterpretation of Table 2. Table 2 presents included systematic reviews; it does not present individual studies included in these reviews.

• Within the reporting of results in the text i.e. when SMD and RR are presented, the authors should present the number of trials and number of participants on which the result is based.

Response: We present this information in Figures 3 to 5. We are concerned that this information might be redundant if we present it in figures and text. We'd be happy to add it to the text, however, if the BMJ Open editors think that presenting such data twice, albeit in two different formats, would benefit readers.

• It is encouraging to see the comment on the issues of publication bias when reporting the results for SGA's. Often these trials rate quite well in terms of risk of bias, but there are a number of bias's outside of e.g. the Cochrane risk of bias tool that impact on pharmacological trials.

Response: We agree with the reviewer. Other meta-biases, such as funding bias, could indeed have an impact on results and should be examined in systematic reviews. In the case of our study, however, which is a review of systematic reviews, identifying and possibly needing to deal with this possible problem would have required us to re-review all included trials for each systematic review to be able to assess funding bias. We added a sentence to the discussion alerting readers that other biases could have an impact on results.

• Some more consistency in terms of reporting results is required; I noted that for some results the strength of evidence is reported but e.g. for non pharmacological interventions it is not.

Response: We added the strength of evidence ratings where they were missing in the text. In

addition, we present them for all comparisons in figures 3-5.

• Page 19, line 17 – typo – should read “bias”.

Response: We fixed the typo.

• Page 23, line14 – typo: decision making. Author should also include a reference with regard to their statement about shared decision making improving treatment adherence etc…
Response: We fixed the typo and added a reference

• I think the other area for future research is around efficacy of non pharmacological interventions?
Response: Excellent point – we added text to the discussion

Reviewer 3

• This paper is heavily based on the previous work of the authors in the same field (for instance, see refs 46, 48 and 52 of the present manuscript). I think it is interesting for clinicians to have a "short" summary of all the previous evidence/papers, but the authors should make clear (since from the introduction) why the present manuscript is different from their previous publications (refs 46, 48 and 52 should probably be quoted in the introduction). I think this article could benefit from a slightly different angle, in order to avoid the replication of findings that have been already presented/discussed in the literature very recently (I suggest a larger time window and a more recent update of the search strategy).

Response: The reviewer is absolutlely correct: some of the findings are from our previous systematic review which, just like other systematic reviews, got picked up by the literature searches. The main goal of the paper, however, was to expand the scope to more than 140 pharmacological and nonpharmacological interventions that our prior review(s) had not addressed. Unfortunately, we did not find much evidence on these interventions, which is probably a finding by itself. Our goal was not to re-package the prior paper but rather to provide a more comprehensive overview on pharmacologic and nonpharmacologic interventions.

Here below a few (mainly minor) comments:

• The conclusion paragraph int he abstract mentioned only non-pharmacological interventions, while the paper is also about pharmacological treatments. Please amend accordingly

Response: We added text to the Conclusion

• The first sentence of the introduction is not relevant, as the article focused on MDD. I would remove it

Response: We agree with the reviewer, this sentence indeed does not add much. We removed it in the revised manuscript.

• Page 6, line 56: probably the acronym PICOTS is useless (it is reported once more in the table only and not in the text)

Response: Thanks for pointing this out. We deleted PICOTS.

• Please clarify why the search is limited between 1 January 2011 and 23 February 2016? If reference 25 is true, the authors should now include papers from 2012! I would not be so strict, but I think the

search should be a bit broader and, if it is not possible to go back before 2011, I would certainly update it up to Jan 2017 (it was carried out 11 months ago, so it would be important that the authors could check whether any relevant papers have been published since Feb last year)

Response: For the revised manuscript, we conducted an updates search until February 2017. We included 4 new or more recent reviews. We chose the 5.5 year cut-off because methods research indicates that after 5.5 years, 50% of systematic reviews are outdated. For the revised manuscript, we do not include any systematic reviews from 2011. Therefore, for simplicity, we would like to keep the 2011 search limit.

• Sorry, I couldn't find in the text, tables or appendices any definition of "systematic review" or "meta-analysis". If I am not mistaken, it would be useful for the readers to know exactly the kind of studies included in the present article (sometimes "narrative reviews" are called "systematic review" in the title, but I'm sure the authors wanted to include only properly carried out systematic reviews)

Response: Our definition of systematic reviews was based on the Cochrane handbook. We added a footnote to Table 1 to clarify the definition.

• Table 1 is not clear to me. Why agomelatine is alone? Is it because of a different mechanism of action? If so, why did you include vortioxetine among other antidepressants? I think an explanation of the rationale for these choices should be presented/discussed in the main text (or in the table itself).

Response: Thanks for pointing this out. We consolidated the classifications and now have agomelatine and vortioxetine within the second-generation antidepressants category.

• Table 1: sorry, what does it mean exactly "Pharmacologic interventions (for comparison with inactive interventions)"?

Response: We deleted “(for comparison with inactive interventions).“

• Table 1: "We did not include combination treatments" and "as an initial monotherapy" are redundant. I would use one or the other

Response: We believe being clear on this point is important because several studies did assess combination therapies of antidepressants and complementary medicines. Reviews on such studies would not have been eligible for our specific purpose.

• Table 1: please provide a clear definition of MDD (which criteria used) and response (which threshold, if any). Did the authors consider change of depression scores based only on validated scale? Did they consider both self-rating and clinician-based scales?

Response: Because our work was a review of systematic reviews, we depended on the definitions that authors of included systematic reviews used. For MDD, sometimes this was based on DSM criteria; often, however, the study authors did not give a definition. Similarly, response to treatment was sometimes defined as an improvement of symptoms of at least 50% but, again, often not clearly defined. We address these issues in the Discussion under limitations. Therefore, MDD and response are “as defined by authors” who probably relied on definitions within RCTs. As for any review of systematic reviews, this approach introduces some vagueness of definitions. This is also the reason that we caution readers to compare effect sizes across interventions.

• Reference 51 should be replaced by a more recent publication with updated results: Furukawa et al.

Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies. Lancet Psychiatry. 2016 Nov;3(11):1059-1066.

Response: We replaced this reference with the more recent one provided by the reviewer.

• Page 21, line 5: Is "risk of adverse events" one of the outcomes? This is not a trivial issue to clarify, not only because the risk of adverse events was not listed as an outcome in the method section of this paper and also because ref 46 included non-randomised evidence (which is in conflict with the methods as stated on page 7, first two lines.

Response: Thanks for pointing this out. “Risk of adverse events” was not an eligible outcome. We do not present any data on this outcome. We deleted “risk of adverse events” in this sentence.

Reference 46 (our previous report, Gartlehner et al.) indeed included non-randomized studies. However, all of the outcomes and data that we present in this review of reviews are based on RCT evidence. Non-randomized studies did not contribute to any of these estimates.

• Page 23, line 6: please change "We believe that our results have important clinical implications." into "We believe that our results MAY have important clinical implications."

Response: We changed the wording.

Reviewer 4

• This study investigate reviews research evidence from more than 140 pharmacologic and on-pharmacologic treatment options for major depressive disorder (MDD) obtained from systematic reviews of randomized controlled trials in adult patients with acute-phase MDD. The review resulted in identification of 15 systematic reviews on 27 comparisons of interest. This review suggested strong evidence supporting the small benefits of second-generation antidepressants but also significant rate of discontinuation due to adverse events. Cognitive behavioral therapy was found to show reliable evidence similar to those of second-generation antidepressants. There was no evidence supporting majority of non-pharmacologic interventions for treating MDD. While this study addresses an important research question, I have some methodological concerns about the pooling of evidence from the include articles.

On pages 10 -11, the authors provided details about meta-analysis of evidence. One feature of meta-analysis of evidence from multiple systematic reviews is the increase layer of sources of variations introduced. Ideally, pooling evidence from multiple systematic reviews leads to multiple sources of variations such as between-systematic review variation, within-systematic review variation, and individual-level study variation. It is not clear if the authors accounted for between-study variations in their analyses and whether the conclusions would have different have they accounted for these sources of variations.

Response: In our review, we did not pool studies across systematic reviews for exactly the reasons that the reviewer outlines above. The only instance for which we recalculated a meta-analysis was when individual RCTs of eligible systematic reviews did not meet our eligibility criteria (e.g., because they used treatment as usual as a control group). We then recalculated the meta-analysis removing ineligible studies.

• It appears the authors used conventional meta-regression methods to pool evidence but use network meta-analysis in others. This is not clearly stated anywhere in the manuscript. In fact, network meta-analysis was only mentioned in the Result section.

Please provide a detailed description of the network meta-analysis employed as part of the statistical analysis section of the manuscript.
The authors used conventional meta-regression approaches and network meta-analytic methods to address the research questions. I would recommend that the author indicate these in the methods section of the abstract.

Response: For all meta-analyses, we used standard methods (DerSimonian&Laird) as employed by the Cochrane Collaboration's RevMan software. We describe this approach in the Methods. We did not conduct network meta-analyses for this review of systematic reviews, we just presented a network graph in Figure 2. Because this figure indeed can cause confusion, we removed it from the revised manuscript.

• As part of the discussions about the limitations of this review, the authors need to highlight the small number of studies (low power, low sample size) as a limitation of the meta-analysis approaches investigated.

Response: Thank for raising this important point. We added text to the limitations.

• On page 22 line 8, the authors stated "Although most of the reviews 9 had few problems in methods, conceivably these authors did miss some RCTs. Conceivably, 10 RCTs are available…". Conceivably used twice, please revise.

Response: We revised these sentences.

VERSION 2 – REVIEW

REVIEWER	Klaus Linde Technical University Munich, Germany
REVIEW RETURNED	22-Mar-2017

GENERAL COMMENTS	The authors adequately addressed my comments

REVIEWER	Dr. Tolulope Sajobi University of Calgary, Calgary, Canada
REVIEW RETURNED	11-Apr-2017

GENERAL COMMENTS	The authors have satisfactorily addressed my comments.